

## Non-destructive shoot biomass evaluation using a handheld NDVI sensor for field-grown staking Yam (*Dioscorea rotundata* Poir.)

Kohtaro Iseki & Ryo Matsumoto

To cite this article: Kohtaro Iseki & Ryo Matsumoto (2018): Non-destructive shoot biomass evaluation using a handheld NDVI sensor for field-grown staking Yam (*Dioscorea rotundata* Poir.), Plant Production Science, DOI: [10.1080/1343943X.2018.1540278](https://doi.org/10.1080/1343943X.2018.1540278)

To link to this article: <https://doi.org/10.1080/1343943X.2018.1540278>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Accepted author version posted online: 23 Oct 2018.  
Published online: 20 Nov 2018.



Submit your article to this journal [↗](#)



Article views: 125



View Crossmark data [↗](#)

# Non-destructive shoot biomass evaluation using a handheld NDVI sensor for field-grown staking Yam (*Dioscorea rotundata* Poir.)

Kohtaro Iseki<sup>a</sup> and Ryo Matsumoto<sup>b</sup>

<sup>a</sup>Japan International Research Center for Agricultural Sciences, Tsukuba, Japan; <sup>b</sup>International Institute of Tropical Agriculture, Ibadan, Nigeria

## ABSTRACT

Crop phenotyping is a key process used to accelerate breeding programs in the era of high-throughput genotyping. However, most rapid phenotyping methods developed to date have focused on major cereals or legumes, and their application to minor crops has been delayed. In this study, we developed a non-destructive method to predict shoot biomass by measuring spectral reflectance in staking yam (*Dioscorea rotundata*). The normalized difference vegetation index (NDVI) was evaluated using a handheld sensor that was vertically scanned from the top to the bottom of a plant alongside the stake. A linear regression model was constructed to predict shoot biomass through Bayesian analysis using NDVI as a parameter. The model well predicted the observed values of shoot biomass, irrespective of the growth stage and genotypes. Conversely, the model tended to underestimate the shoot biomass when the actual shoot biomass exceeded 150 g plant<sup>-1</sup>; this was compensated for when the parameter green area, calculated from plant image, was included in the model. This method reduced the time, cost, effort, and field space needed for shoot biomass evaluation compared with that needed for the sampling method, enabling shoot biomass phenotyping for a large population of plants. A total of 210 cross-populated plants were evaluated, and a correlation analysis was performed between the predicted shoot biomass and tuber yield. In addition to the prediction of tuber yield, this method could also be applied for the evaluation of crop models and stress tolerance, as well as for genetic analyses.

## ARTICLE HISTORY

Received 4 July 2018  
Revised 10 September 2018  
Accepted 18 October 2018

## KEYWORDS

NDVI; Yam; *Dioscorea rotundata*; phenotyping; shoot biomass; Bayesian analysis

## 1. Introduction

The time, cost, and effort associated with field phenotyping are the restrictive factors for breeding programs, where delays in phenotyping have been outstanding with a recent great improvement in genotyping technologies (Araus & Cairns, 2014; White et al., 2012). To date, the remote-sensing methods have been developed to evaluate agronomic phenotypes, such as shoot biomass, plant nutrient status, and abiotic stress tolerance (Knyazikhin et al., 2013; Maes & Steppe, 2012; Yi, Xiang, Ming, Wang, & Zhao, 2013). These methods can evaluate large crop populations within a feasible time frame and help to establish a standard phenotyping protocol in breeding programs. However, most high-throughput phenotyping methods use expensive optical sensors and require specialist knowledge for data analysis. In addition, the application of remote sensing for plant phenotyping has focused mainly on major crops and has been delayed in minor crops.

Yam is a minor, but important, tuber crop grown widely in temperate to tropical regions. The total yam-cultivated area is 7.5 million hectares, with an annual

harvest of 65.9 million tons of tubers (FAOSTAT; <http://www.fao.org/faostat/en>). West Africa is the largest yam-cultivating area, accounting for more than 94% of the world's yam production, with *Dioscorea rotundata* being the major species used for local consumption. Conversely, the development of varieties in this region has not benefitted from modern breeding, and the average tuber yield has been stagnant for more than five decades. One reason for this slow breeding progress is the lack of phenotypic information on genetic resources, although the number of yam accessions in West Africa is larger than that in the other regions.

The technical challenges associated with yam breeding and selection include the long breeding cycle, extremely low multiplication ratio of propagules, and existence of a juvenile phase during the seminal and early clonal stages of selection (Asiedu & Sartie, 2010). One of the difficulties of the field-based yam phenotyping is the vine plant type. To avoid interference between neighboring plants, staking has been widely adopted in experimental trials (Law-Ogbomo & Remison, 2008; Otoo & Asiedu, 2009; Suja, Nayar, & Sreekumar, 2005). However, staking makes it difficult

to apply common remote-sensing techniques for above-ground traits, because most techniques were developed for crops with an erect plant shape and a relatively low plant height. In staked vine plants, leaves are positioned in a vertical direction, which is unsuitable for the horizontal scanning of spectral reflectance from above the plants. In addition, the long growth period, low planting density, and vegetative propagation system requires space, time, and effort for cultivation, and thus, the number of genetic resources has been restricted in previous studies analyzing growth (Diby et al., 2011; Hgaza, Diby, Assa, & Ake, 2010; Macros, Cornet, Bussi re, & Sierra, 2011).

The purpose of this study is to develop a rapid and non-destructive method of evaluating shoot biomass of staking yam using an optical sensor. Normalized difference vegetation index (NDVI) has been used to evaluate shoot biomass in various crops, including rice, barley, and maize (Ali, Thind, Sharma, & Singh, 2014; Elsayed, Rischbeck, & Schmidhalter, 2015; Teal et al., 2006). NDVI is calculated from the equation  $NDVI = (IR - R) / (IR + R)$ , where IR and R are the canopy reflectances of the near infra-red and red spectra, respectively. This is dependent on the notion that the red spectra show low reflectance because of the absorption by chlorophyll, while the infra-red spectra show high reflectance because of internal leaf scattering and no absorption (Knipling, 1970). The NDVI represents the fractional vegetation cover and leaf area index (Carlson & Ripley, 1997), showing the value of 0.7–0.8 for a fully covered canopy. A low-cost handheld optical sensor for measuring NDVI was applied to staking yam, with some modification of the measurement procedure. A regression model was constructed to predict shoot biomass through Bayesian data analysis.

## 2. Materials and methods

### 2.1. Plant materials and growth conditions

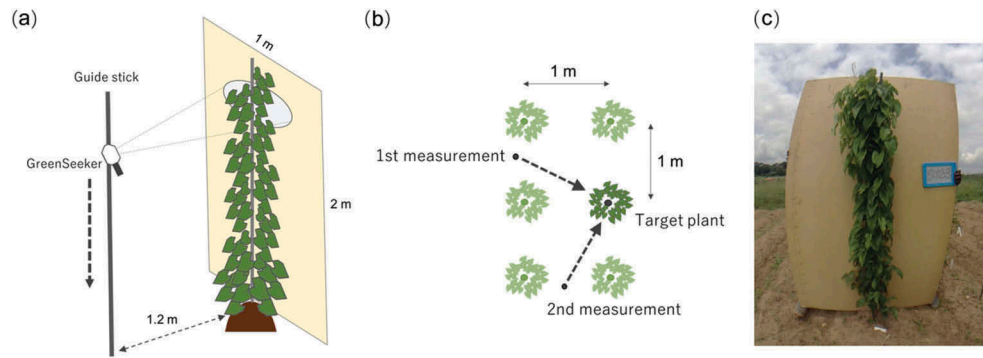
The field experiments were conducted during the yam growing season (May–December) in 2016 and 2017 at the International Institute of Tropical Agriculture (IITA) in Ibadan, Nigeria (7°29'N, 3°54'E). The total precipitation, average maximum temperature, and average minimum temperature during the growth periods were 1166 mm, 22.5°C, and 30.6°C in 2016 and 1309 mm, 22.7°C, and 30.4°C in 2017, respectively. The soil was classified as sandy loam with moderate acidity (Dare, Abaidoo, Fagbola, & Asiedu, 2008). A *D. rotundata* accession (TDr 95/01932) from IITA was used for planting. To obtain differences in shoot biomass, we used different sizes of tuber pieces for planting, considering

that shoot biomass was highly affected by seed tuber size (Cornet, Sierra, Tourn bize, Gabrielle, & Lewis, 2016). Tuber pieces (setts) of 50, 100, and 200 g were cut from the middle of the tubers. To ensure equivalent growth conditions, the planting was performed in late-May in a plastic pot with a 12 cm diameter and 10 cm height; thereafter, only well-sprouted plants were transplanted to the experimental field in early-June. For each sett size, 12 plants were transplanted at a spacing of 1 m × 1 m on the plots of 3 m × 4 m, arranged in a randomized block design. Three and five plot replicates were used in 2016 and 2017, respectively. Stakes were installed at the same time as transplanting. Weeding was conducted every 2 weeks using a hand hoe. Fertilizers were not applied.

### 2.2. Measurements

A diagram of the procedure for NDVI measurements using a handheld optical sensor in staking yam (GreenSeeker, Nikon Trimble, Tokyo, Japan) is shown in Figure 1(a). The size of the device was 9 cm (width) × 27 cm (length), and the weight was 310 g, making it suitable for operation with only one hand. The GreenSeeker sensor recorded the spectral reflectance of an elliptical area with a 50-cm long axis from a distance of 1.2 m, which is the maximum measurable range of the device. For each plant, a line-scan from top to bottom was performed vertically with a guide stick standing at 1.2 m from the target plant, and the NDVI was automatically calculated as the average value during scanning. The scanning was carefully performed for 30 s per measurement, keeping the scan speed constant from top to bottom. To eliminate the background noise of the reflectance, a board of 1 m × 2 m was set behind the plant. The measurement was performed twice per plant from different directions (Figure 1(b)). The plant height was measured visually at the same time as NDVI was measured.

The above measurements were conducted two and four times during the growing season in 2016 and 2017, respectively. The measurements were made in September and November 2016, corresponding to three and five months from transplanting, and in July, August, October, and December 2017, corresponding to two, three, five, and seven months from transplanting, respectively. The NDVI and plant height were measured for two plants per plot. All shoot parts were sampled after the measurements, and the shoot dry weight was determined after oven-drying at 80°C for 2 days. In the data analysis, shoot biomass was represented as the shoot dry weight.



**Figure 1.** Normalized difference vegetation index (NDVI) measurement procedure for staking yam and a plant image taken by a wide-angle lens. (a) NDVI is scanned from top to bottom alongside the target plant. The guide stick stands at a distance of 1.2 m from the plant. (b) Positions of NDVI measurement in the yam population as shown from the top. NDVI was measured twice per plant from different directions. (c) Plant image was taken with a board behind the plant from the same position as that of NDVI measurement.

In addition to the NDVI and plant height, plant images were also used as a model parameter. Plant images were obtained at the same time as NDVI was measured, on August 2017. A wide-angle lens (PL-WD02, Aukey, Shenzhen, China) was attached to the camera installed on an iPad mini (Apple, Cupertino, USA), and the camera direction was set vertical to the board behind the plant. Whole plant images were obtained from a middle point of the guide stick for NDVI measurement, approximately 1 m above the ground, standing 1.2 m from the target plant (Figure 1(c)). The projected green area in the image was calculated using the image analysis software ImageJ1 version 1.50i (National Institute of Health; <http://rsb.info.nih.gov/ij/>). The board width was used as the reference length for each image. The time of taking images and sunlight direction did not affect image analysis. Barrel distortion due to the use of a wide-angle lens was not corrected, and thus the calculated green area was different from the actual value.

## 2.3. Model construction

### 2.3.1 Parameter selection

The shoot biomass was estimated from a linear regression model through Bayesian analysis. The data from August 2017, with  $n = 10$  for each sett size, were used for the models. For input parameters, we first tested all combinations among NDVI, plant height, and green area, and then the constructed models (Equations 1–7) were compared using the widely applicable information criterion (WAIC) (Watanabe, 2010). To evaluate parameter contribution in the prediction of shoot dry weight, all input variables were standardized with a mean of 0 and a standard deviation of 1 before analysis.

$$\text{basal biomass} = a \times \text{NDVI} + \text{constant} \quad (1)$$

$$\text{basal biomass} = b \times \text{Green area} + \text{constant} \quad (2)$$

$$\text{basal biomass} = c \times \text{Plant height} + \text{constant} \quad (3)$$

$$\text{basal biomass} = a \times \text{NDVI} + c \times \text{Plant height} + \text{constant} \quad (4)$$

$$\text{basal biomass} = a \times \text{NDVI} + b \times \text{Green area} + \text{constant} \quad (5)$$

$$\text{basal biomass} = b \times \text{Green area} + c \times \text{Plant height} + \text{constant} \quad (6)$$

$$\text{basal biomass} = a \times \text{NDVI} + b \times \text{Green area} + c \times \text{Plant height} + \text{constant} \quad (7)$$

where  $a$ ,  $b$ , and  $c$  are the coefficients of the parameters of NDVI, green area, and plant height, respectively. The shoot biomass was considered to follow a normal distribution with an average basal biomass and a standard deviation of  $\sigma$  (Equation 8).

$$\text{Shoot biomass} \sim \text{Normal}(\text{basal biomass}, \sigma) \quad (8)$$

Where,  $a$ ,  $b$ ,  $c$ ,  $\text{constant}$ , and  $\sigma$  are estimated from the posterior distribution generated by the Markov chain Monte Carlo (MCMC) method. The MCMC algorithm was set at 3,000 steps for iteration and 500 steps for warm-up; the number of chains was four, and the total sample size was 10,000. The convergence was confirmed by visualization of a trace plot and 'R hat' (potential scale reduction factor on split chains). The Bayesian analysis was performed using the statistical software R version 3.4.1 with the package 'rstan'.

### 2.3.2. Effect of growth stage on the model

A model of Equation (1) was used for further analysis. To confirm the robustness of the model for each growth stage of yam, we compared the coefficients of the models constructed using the data from different growth stages (the number of months after transplanting). The data from each growth stage were  $n = 18$  and  $30$ , including each sett size, in 2016 and 2017, respectively. The variables were not standardized, and the raw values were used for model construction. The hierarchical Bayesian model was applied as follows:

$$\text{basal biomass} = a[k] \times \text{NDVI} + \text{constant}[k] \quad (9)$$

$$x[k] \sim \text{Normal}(x', \sigma_x) \quad (10)$$

where  $a[k]$  and  $\text{constant}[k]$  were different for each growth stage. The number of  $k$  was six, corresponding to September and November in 2016, and July, August, October, and December in 2017. Here, we assumed that the differences in the coefficients of the models among the growth stages were followed by a normal distribution. In Equation (10),  $x[k]$  represents each of  $a[k]$  and  $\text{constant}[k]$  in Equation (9), and  $x'$  and  $\sigma_x$  represent the mean and standard deviation of the normal distribution of  $x[k]$ . Equation (8) was used to predict the distribution of shoot biomass. The posterior distributions of all coefficients were generated using the same method as describe above for MCMC.

### 2.4. Model validation

We constructed a final model covering different growth stages. This model was applied to the new data set for shoot biomass and NDVI recorded in different field trials. Twenty-eight *D. rotundata* accessions presenting large diversity in shoot biomass and tuber yield were selected from the Genetic Resources Center, IITA. A sett of size 100 g was planted in mid-May 2017, and well-sprouted plants were transplanted to the experimental field of IITA in early-June. Six plants were transplanted at a spacing of 1 m  $\times$  1 m on the plots of 2 m  $\times$  3 m, arranged in a randomized block design with three replicates. Stakes were installed at the same time as transplanting. Weeding was conducted every 2 weeks using a hand hoe. Fertilizers were not applied.

The NDVI was measured for a plant from each replicate on September 2017, three months after transplanting. Total shoots were sampled from the plants used for NDVI measurement on the same day. The shoot dry weight was measured after oven-drying at 80°C for 2 days. The collected data were used for constructing the final model, and the validity was analyzed by performing a correlation analysis using the statistical software R. The data used for the validation were  $n = 28$  with three replicates and total  $n = 84$ .

### 2.5 Tuber yield expectation from predicted shoot biomass

After model validation, the final model was applied to estimate the shoot biomass of the *D. rotundata* cross population comprising 211 plants. The  $F_1$  population generated from a cross between TDr 97/00777 and TDr 04–219 was grown in the experimental field of IITA. A sett of size 100 g was planted in the plastic pot in mid-May in 2016 and 2017, followed by the transplanting of well-sprouted plants in early-June. The field design included one plant for each of the three replications in a randomized block design with a spacing of 1 m  $\times$  1 m. The stakes were installed at the same time as transplanting. Weeding was conducted every 2 weeks using a hand hoe. Fertilizers were not applied.

In October in 2016 and 2017, five months after transplanting, the NDVI was measured for all plants. Subsequently, tubers were harvested at full maturity in December. To analyze the relationship between tuber yield and predicted shoot biomass from the final model, a correlation analysis was performed using the same statistical software.

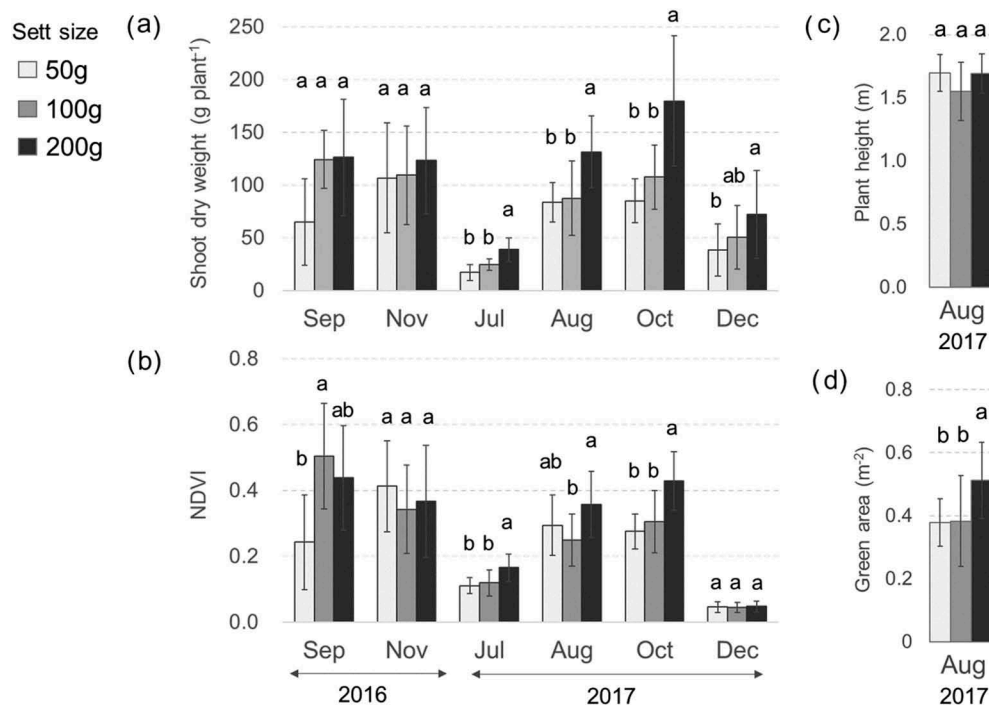
## 3. Results

### 3.1. Data for model construction

The parameters used for the model construction are shown in Figure 2. In 2016, the average shoot biomass of 50-g sett plants was less than that of 100 and 200 g in September, although the differences were not statistically significant (Figure 2(a)). Conversely, the shoot biomass was similar among plants in November. In 2017, the shoot biomass of 200-g sett plants was higher than that of 50- and 100-g sett plants, irrespective of growth stage. The shoot biomass increased from July until October and then decreased. The shoot biomass used for model construction ranged from 5.0 g plant<sup>-1</sup> in 50-g sett plants in July 2017 to 286.5 g plant<sup>-1</sup> in 200-g sett plants in October 2017.

The differences in NDVI among the plants followed the same tendency as that in the shoot biomass, but no differences were observed in December 2017 (Figure 2(b)). The NDVI ranged from 0.03 in 50-g sett plants in December 2017 to 0.70 in 100-g sett plants in September 2016. The plant height and green area were recorded only in August 2017 for parameter selection. No significant differences were observed in plant height among different sett sizes (Figure 2(c)). The minimum plant height was 1.0 m in 100-g sett plants, while the maximum value was restricted to 2.0 m due to the height of stakes. The green area was higher in 200-g sett plants, but did not show significant differences between the plants with 50- and 100-g sett sizes (Figure 2(d)).





**Figure 2.** Summary of the data used for model construction. (a) Shoot dry weight, (b) Normalized difference vegetation index (NDVI), (c) Plant height, and (d) Green area. Values are mean  $\pm$  standard deviation for plant replicates of  $n = 6$  in 2016 and  $n = 10$  in 2017. Bars with the same letters are not significantly different at  $p < 0.05$ .

### 3.2. Parameter selection

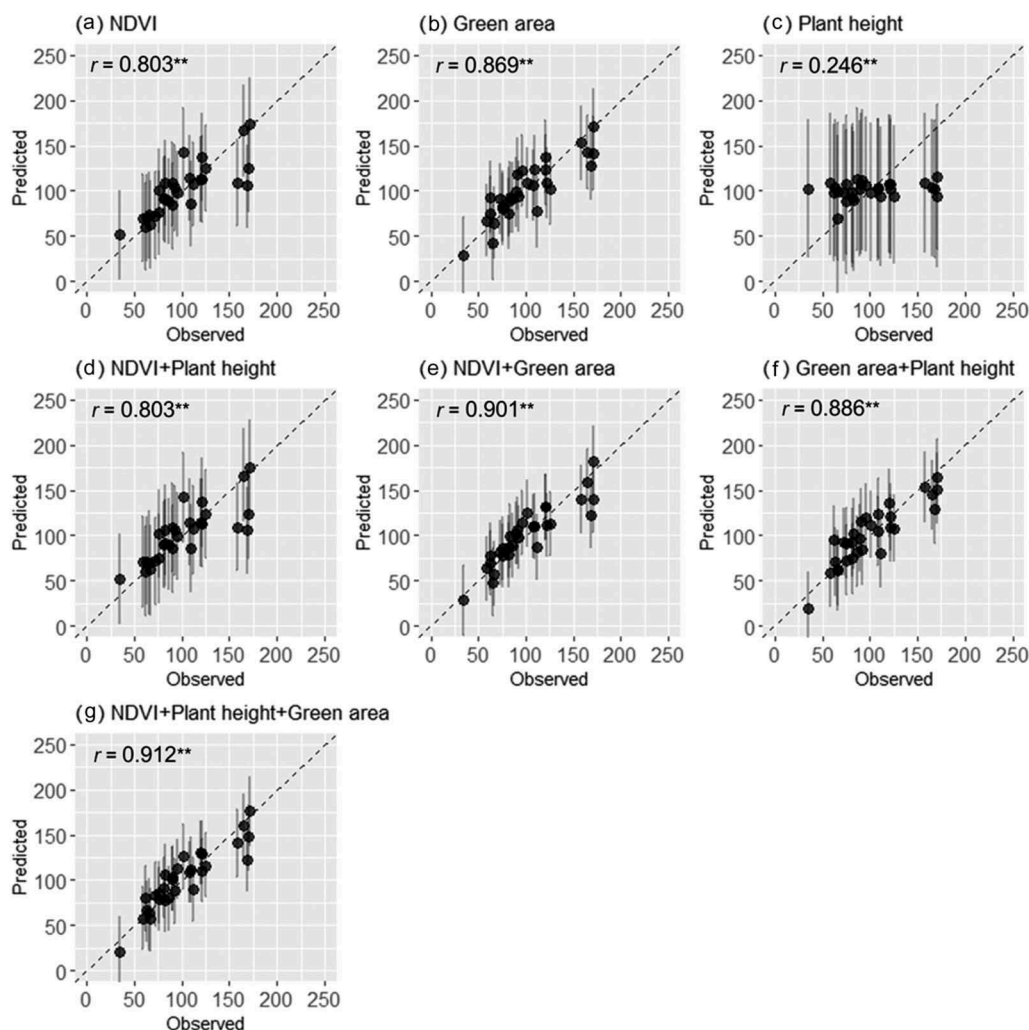
The predictions of shoot dry weight using the constructed models from Equations (1–7) are shown in Figure 3. The most well-predicted model was Equation (7), which used all three parameters; it showed the highest correlation coefficient and the lowest WAIC (Table 1). The models with green area (Figure 3(b,e–g)) predicted better than the models without green area (Figure 3(a,c,d)). In Equations 5 and 7, the higher values of the coefficient with green area than those with NDVI indicate the larger contribution of the green area for shoot dry weight prediction (Table 1). The plant height did not contribute to the prediction of shoot biomass, and Equation (3) showed the lowest correlation coefficient and the highest WAIC.

The model accuracy for plants with low shoot biomass (less than  $150 \text{ g plant}^{-1}$ ) was almost similar among all models, except for Equation (3), while an underestimation was observed (Figure 3(a,d)) for the plants with a high relative shoot biomass (more than  $150 \text{ g plant}^{-1}$ ) in the models without a green area. However, even in these models, some plants with a shoot biomass exceeding  $150 \text{ g plant}^{-1}$  presented good predictions.

### 3.3. Effects of growth stage on the model

To compare the models constructed for each growth stage, the coefficient distributions are shown in Figure 4 (a). The distributions of coefficient  $a$  were similar, except in December 2017, when the values were higher than those at the other growth stages. The differences in the *constant* were negligible among the growth stages. The constructed models for each growth stage are shown in Figure 4(b). Throughout the growth stages, the shoot biomass was well predicted in most plants, with 95% prediction intervals including an equal line between the predicted and observed values. Conversely, the shoot biomass was severely underestimated in the plants with relatively high shoot biomass in October 2017.

We constructed a final model integrating several growth stages, except for that in December 2017. A summary of the posterior distributions of the coefficients are shown in Table 2. All coefficients converged and showed that the  $R^2$  was  $< 1.1$ . This model tended to underestimate the shoot dry weight when this exceeded  $150 \text{ g plant}^{-1}$ ; this occurred mostly in the 200-g sett plants.



**Figure 3.** Correlations between the predicted and observed values of shoot biomass. Models (a–g) correspond to Equations (1) to (7), respectively. The dashed line is an equal line between the predicted and observed values. Points and bars represent the mean and 95% interval of the predicted distribution, respectively. All the units are g plant<sup>-1</sup>. In each panel, the correlation coefficient was statistically significant at  $p < 0.05$ .

### 3.4. Model validation and the relationship between shoot biomass and tuber yield

The final model was validated using a new data set comprising 28 yam accessions. The predicted values were well correlated with the observed values such that all 95% prediction intervals included an equal line between the predicted and observed values (Figure 5). The observed values ranged from 16 to 164 g plant<sup>-1</sup>, and no severe underestimation was observed.

Furthermore, the final model was applied to the F<sub>1</sub> cross population of 210 plants. Figure 6 shows the relationships between the predicted shoot biomass and observed tuber yield. The predicted shoot biomass varied from 34.8 to 146.7 g plant<sup>-1</sup> in 2016 and from 57.3 to 147.2 g plant<sup>-1</sup> in 2017. Although the average shoot biomass was similar in the two years, the tuber

yield was relatively low in 2017. In both years, the correlation coefficient exceeded 0.6. The correlation coefficient of the tuber yield between the two years was 0.28, and a similar value ( $r = 0.29$ ) was observed for the correlation coefficient of the predicted shoot biomass. Figure 6 also shows the existence of a large variation in tuber yield, even among the accessions with the same level of shoot biomass.

## 4. Discussion

In this study, a rapid and easy method was developed for the evaluation of shoot biomass in staking yam. This method can reduce the time and cost required to evaluate shoot biomass compared with the destructive sampling method. It could evaluate more than 200 accessions within 4 h, including the preparation time, although it

**Table 1.** Means of posterior distribution of the coefficients and widely applicable information criterion (WAIC).

Model	Coefficient				WAIC
	<i>a</i> (NDVI)	<i>b</i> (Green area)	<i>c</i> (Plant height)	constant	
(1) $a \times \text{NDVI} + \text{constant}$	29.4	-	-	100.7	276.1
(2) $b \times \text{Green area} + \text{constant}$	-	31.8	-	100.8	264.3
(3) $c \times \text{Height} + \text{constant}$	-	-	8.9	100.8	304.2
(4) $a \times \text{NDVI} + c \times \text{Plant height} + \text{constant}$	29.2	-	1.3	100.7	277.6
(5) $a \times \text{NDVI} + b \times \text{Green area} + \text{constant}$	13	22.2	-	100.8	259.3
(6) $b \times \text{Green area} + c \times \text{Plant height} + \text{constant}$	-	35	-7.1	100.9	261.7
(7) $a \times \text{NDVI} + b \times \text{Green area} + c \times \text{Plant height} + \text{constant}$	12.2	25.4	-5.9	100.7	258.0

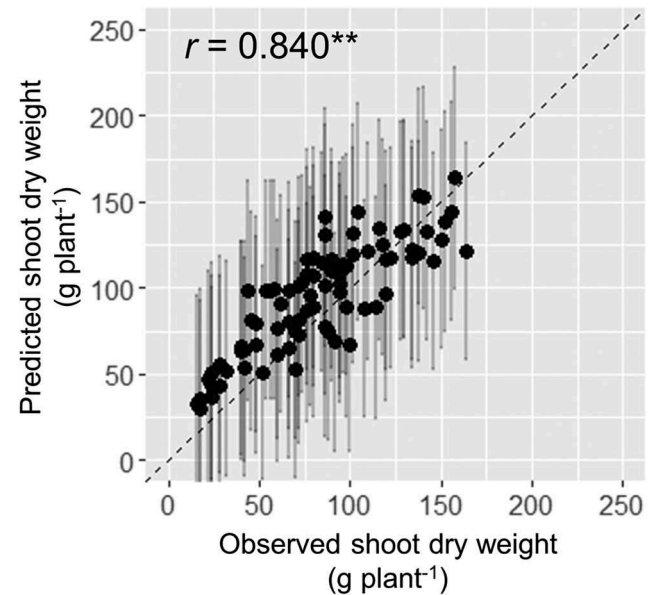
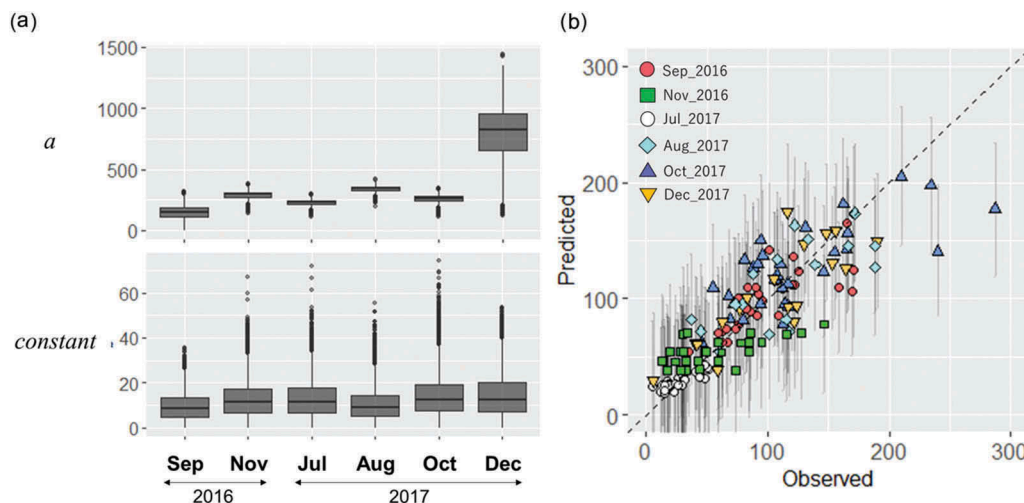
Means of the 10,000 Markov chain Monte Carlo (MCMC) samples for each of the coefficients are shown. All variables were standardized before analysis.

might take at least 3 days through the ordinary method of shoot sampling, oven drying, and weighing. In addition, owing to the non-destructive measurement of NDVI, evaluators can pursue the growth of the same plant from planting to harvest. This also reduces the size of the trial field needed for plant replication during time-course destructive sampling. Instead, for destructive sampling, we can increase the number of plant replicates, which helps to avoid unreliable biomass data resulting from high inter-plant variability of yam (Cornet, Sierra, Tournebize, & Ney, 2014). Because the price of the GreenSeeker sensor is less than 1,000 USD, which is much lower than that of the other optical sensors used for measuring spectral reflectance, such as multi-spectral and hyper spectral cameras, it can be easily introduced into yam breeding programs worldwide. Additionally, no

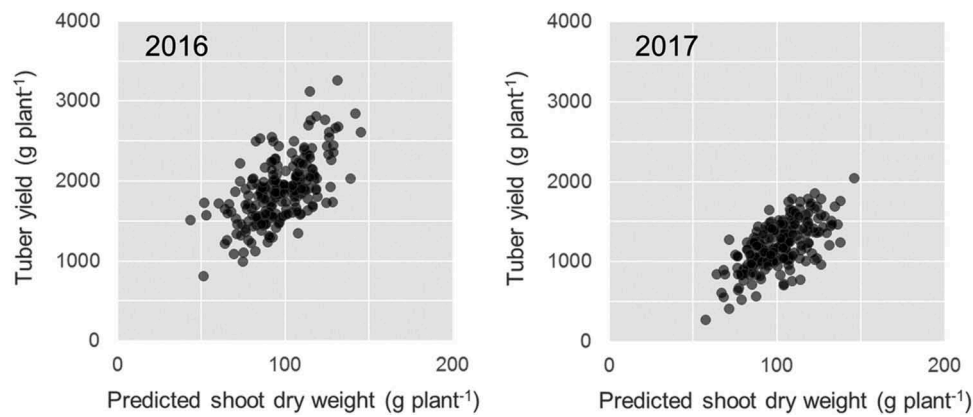
**Table 2.** Summary of the posterior distributions in the final model.

Coefficient	Parameter	Mean	SD	95% prediction interval	
				2.5%	97.5%
<i>A</i>	NDVI	297.6	13.2	269.5	320.4
Constant		4.7	3.7	0.2	13.4
$\sigma$		32.0	2.0	28.3	36.2

Mean, standard deviation (SD), and quantile values were generated from 10,000 Markov chain Monte Carlo (MCMC) samples for each of the coefficients. In the final model, the predicted mean shoot biomass is expressed as  $\text{basal biomass} = 297.6 \times \text{NDVI} + 4.7$ . The variables were not standardized for the model construction, and the raw values were available for shoot biomass prediction.

**Figure 5.** Model validation using 28 accessions of yam. Data from  $n = 28$  with three replicates were used. The correlation between the predicted and observed values was statistically significant at  $p < 0.05$ . Points and bars are the mean and 95% interval of the predicted distribution, respectively.**Figure 4.** Comparison of the models constructed for different growth stages in 2016 and 2017. (a) Posterior distributions of the coefficients. (b) Correlations between the predicted and observed values of shoot biomass. The box plots of the coefficients were generated from 10,000 Markov chain Monte Carlo (MCMC) samples for each growth stage. Points and bars in the scatter plot are the mean and 95% interval of the predicted distribution, respectively.



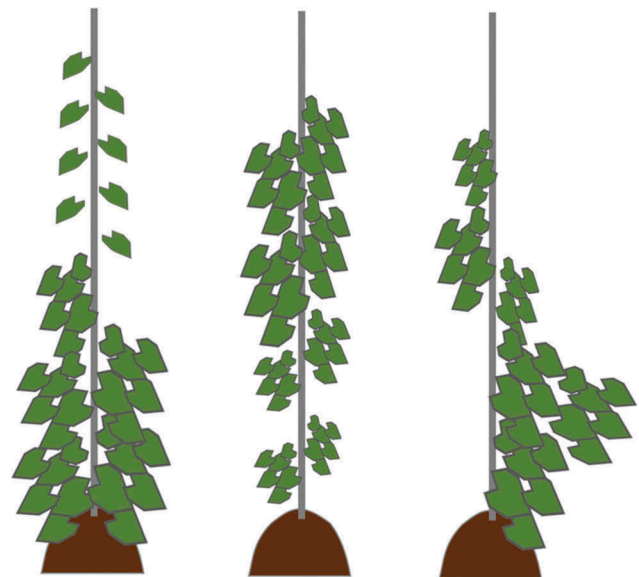


**Figure 6.** Relationships between the predicted shoot biomass and observed tuber yield in the cross population. Correlations were statistically significant at  $p < 0.05$  in both years. Data represent the means of three replicates.

specialist knowledge is needed when using this method to analyze the spectral reflectance data and to record the value of NDVI.

We finally selected a model lacking a green area, but showing better prediction; this implied a relatively low WAIC value, which was found in the models with green area (Table 1). This is because the evaluators would prefer a parameter that needs no post-measurement analysis. The image analysis to detect the projected green area is relatively time-consuming and unfavorable for a large population. The models without green area showed a good prediction, which was underestimated when the shoot biomass exceeded  $150 \text{ g plant}^{-1}$ . This is a common phenomenon with shoot biomass estimation using NDVI (Teal et al., 2006) because it is calculated from the spectral reflectance of a plant surface, which cannot be assessed inside the leaf layer. Therefore, an NDVI value plateaus after the scanning area is covered by the leaves. In this case, the use of the green area calculated from the plant image well compensated the predictions (Figure 3), especially when the plants with large shoot biomass, more than  $200 \text{ g plant}^{-1}$ , were used for evaluation. However, in this study, the shoot biomass rarely reached  $200 \text{ g plant}^{-1}$ , even in the plants with 200-g setts (Figures 2, 5, 6). This is consistent with the findings from the previous studies on *D. rotundata* (Chowdhury, 1998; Suja, 2005; Suja et al., 2005). Thus, considering the time taken for evaluation, a model without a green area is suitable.

Uneven leaf distribution can also prevent good prediction of shoot biomass, as shown in Figure 7. Such plants were observed during the late period of yam growth and caused a failure of NDVI measurement. Even in this case, plant images might improve the predictions, and the other methods of scanning, such as point scans instead of line scans, might also be



**Figure 7.** Plant shapes of staking yam not suitable for shoot biomass prediction using the normalized difference vegetation index (NDVI). Left and middle: leaves are dense in the top or bottom part. Right: leaves are inclined towards one side.

applicable. Moreover, the images from above the plant might improve the predictions by adding information on the vertically projected area. Further studies are needed to improve the method in order to apply it to a wide variety of genotypes and growth conditions.

We had expected that the plant height can detect differences in shoot biomass between the plants that have different heights and the same NDVI values, as obtained from the line scan. However, the plant height did not contribute to the prediction of shoot biomass as the coefficients were lower than the other parameters, and the same WAIC values were obtained for the models with and without plant height (Table 1). This might be because the plant height was not varied among the

plants at the same growth stage, and the leaves constituted more than 60% of total shoot biomass (Chowdhury, 1998; Suja, 2005). Therefore, the measurement of NDVI or green area is critical for the precise prediction of shoot biomass. On the other hand, the plant height is thought to be effective when dwarf-type accessions of yam having high NDVI and low plant height are included in the target population (Abraham, Nair, & Sreekumari, 1989).

The robustness of the prediction of a model was confirmed only by using the NDVI as an input parameter. This final model well predicted the shoot biomass, irrespective of the yam accessions and growth stages studied, although the model was predicted before plant senescence (Figure 4). In December, a yam plant shows shoot senescence with changes in the characteristics of spectral reflectance. This explains the extremely high coefficient of NDVI during this period; therefore, we removed the data of December 2017 from the final model. This final model was applied to shoot biomass prediction of the cross population, and a good correlation was found between shoot biomass with tuber yield (Figure 6). This indicates that the information on shoot biomass is also applicable for tuber yield prediction as shown in the past studies using a restricted number (less than 10) of yam genotypes (Chowdhury, 1998; Diby et al., 2011; Macros et al., 2011). The good relationships observed between shoot biomass and tuber yield in both years also indicate the reliability of this method.

This non-destructive method enabled the determination of a relationship among more than 200 yam plants in less than half a day. This method also led to the detection of a large genotypic variation in tuber yield among plants with similar shoot biomass. One of the possible reasons could be the differences in the date of tuber initiation, which is known to be photoperiod-sensitive (Vaillant, Bade, & Constant, 2005). The early tuber initiation enables a long period of tuber enlargement, causing high tuber yield. Tuber yield was also largely different between the two years, although the shoot biomass was almost similar. The relatively low tuber yield in 2017 implied that tuber formation was restricted to the period after the completion of shoot biomass growth. This might be because of the small precipitation during tuber formation from September to November in 2017, although the yearly total precipitation was almost the same in the two years. These findings imply that several factors independent of the shoot biomass affect tuber formation, although shoot biomass is intrinsically important for the final yield.

The information on shoot biomass is widely available for growth analysis and for the evaluation of stress tolerance and crop models (Cornet, Sierra, &

Tournebize, 2015; Srivastava & Gaiser, 2010; Srivastava, Gaiser, Paeth, & Ewert, 2012). In addition, the rapid phenotyping method fulfills the phenotypic information demands of a large number of genetic resources. The full genome sequence of yam has been determined, and more than 20,000 single nucleotide polymorphisms (SNPs) are available for genetic analysis (Tamiru et al., 2017). In combination with genetic tools, shoot biomass evaluation using NDVI will serve as a key method for accelerating yam breeding programs, as shown in major crops (Edae et al., 2013; Prasad et al., 2007).

## Acknowledgments

A part of the data used in this study was generated under the IITA-MAFF yam project of the International Institute of Tropical Agriculture (IITA), supported by the Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Abraham, K., Nair, S. G., & Sreekumari, M. T. (1989). White yam, *Dioscorea rotundata* Poir., dwarf type. *Tropical Agriculture*, 66, 184–186.
- Ali, A. M., Thind, H. S., Sharma, S., & Singh, V. (2014). Prediction of dry direct-seeded rice yields using chlorophyll meter, leaf color chart and GreenSeeker optical sensor in north-western India. *Field Crops Research*, 161, 11–15.
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, 19, 52–61.
- Asiedu, R., & Sartie, A. (2010). Crops that feed World 1. Yams. *Food Security*, 2, 305–315.
- Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, 62, 241–252.
- Chowdhury, S. R. (1998). Characteristics of dry matter production and partitioning in edible yam (*Dioscorea* spp.) cultivars. *Tropical Agriculture*, 75, 434–439.
- Cornet, D., Sierra, J., & Tournebize, R. (2015). Assessing allometric models to predict vegetation growth of yams in different environments. *Agronomy Journal*, 107, 241–248.
- Cornet, D., Sierra, J., Tournebize, R., Gabrielle, B., & Lewis, F. I. (2016). Bayesian network modeling of early growth stages explains yam interplant yield variability and allows for agronomic improvements in West Africa. *European Journal of Agronomy*, 75, 80–88.
- Cornet, D., Sierra, J., Tournebize, R., & Ney, B. (2014). Yams (*Dioscorea* spp.) plant size hierarchy and yield variability: Emergence time is critical. *European Journal of Agronomy*, 55, 100–107.
- Dare, M. O., Abaidoo, R. C., Fagbola, O., & Asiedu, R. (2008). Genetic variation and genotypic × environment interaction in yams

- (*Dioscorea* spp.) for root colonization by arbuscular mycorrhiza. *Journal of Food, Agriculture & Environment*, 6, 227–233.
- Diby, L. N., Hgaza, T. B., Tié, T. B., Assa, A., Carsky, R., Girardin, O., ... Frossard, E. (2011). How does soil fertility affect yam growth? *Acta Agriculturae Scandinavica, Section B -Soil & Plant Science*, 61, 448–457.
- Edae, E. A., Byrne, P. F., Manmathan, H., Haley, S. D., Moragues, M., Lopes, M. S., & Reynolds, M. P. (2013). Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. *The Plant Genome*, 6. doi:10.3835/plantgenome2013.04.0010
- Elsayed, S., Rischbeck, P., & Schmidhalter, U. (2015). Comparing the performance of active and passive reflectance sensors to assess the normalized relative canopy temperature and grain yield of drought-stressed barley cultivars. *Field Crops Research*, 177, 148–160.
- Hgaza, V. K., Diby, L. N., Assa, A., & Ake, S. (2010). How fertilization affects yam (*Dioscorea alata* L.) growth and tuber yield across the years. *African Journal of Plant Science*, 4, 53–60.
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1, 155–159.
- Knyazikhin, Y., Schull, M. A., Stenberg, P., Mottus, M., Rautiainen, M., Yang, Y., ... Myneni, R. B. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E185–E192.
- Law-Ogbomo, K. E., & Remison, S. U. (2008). Growth and yield of white guinea yam (*Dioscorea rotundata* Poir.) influenced by NPK fertilization on a forest site in Nigeria. *Journal of Tropical Agriculture*, 46, 21–24.
- Macros, J., Cornet, D., Bussière, F., & Sierra, J. (2011). Water yam (*Dioscorea alata* L.) growth and yield as affected by the planting date: Experiment and modelling. *European Journal of Agronomy*, 34, 247–256.
- Maes, W. H., & Steppe, K. (2012). Estimating evapotranspiration and drought stress with ground-based thermal remote sensing in agriculture: A review. *Journal of Experimental Botany*, 63, 695–709.
- Otoo, E., & Asiedu, R. (2009). GGE biplot analysis of *Dioscorea rotundata* cultivar Tela genotypes in Ghana. *Journal of Food, Agriculture & Environment*, 7, 134–139.
- Prasad, B., Carver, B. F., Stone, M. L., Babar, M. A., Raun, W. R., & Klatt, A. R. (2007). Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under Great Plains conditions. *Crop Science*, 47, 1426–1440.
- Srivastava, A. K., & Gaiser, T. (2010). Simulating biomass accumulation and yield of yam (*Dioscorea alata*) in the upper Ouémé Basin (Benin Republic)- I. Compilation of physiological parameters and calibration at the field scale. *Field Crops Research*, 116, 23–29.
- Srivastava, A. K., Gaiser, T., Paeth, H., & Ewert, F. (2012). The impact of climate change on yam (*Dioscorea alata*) yield in the savanna zone of West Africa. *Agriculture, Ecosystems and Environment*, 153, 57–64.
- Suja, G. (2005). Impact of nutrient management on biomass production and growth indices of white yam (*Dioscorea rotundata* Poir.) intercropped in a coconut plantation in South India. *Tropical Agriculture*, 82, 173–181.
- Suja, G., Nayar, T. V. R., & Sreekumar, J. (2005). Growth analysis of *Dioscorea* species. *Tropical Agriculture*, 82, 164–172.
- Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., ... Terauchi, R. (2017). Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biology*, 15, 86.
- Teal, R. K., Tubana, B., Girma, K., Freeman, K. W., Arnall, D. B., Walsh, O., & Raun, W. R. (2006). In-season prediction of corn grain yield potential using normalized difference vegetation index. *Agronomy Journal*, 98, 1488–1494.
- Vaillant, V., Bade, P., & Constant, C. (2005). Photoperiod affects the growth and development of yam plantlets obtained by in vitro propagation. *Biologia Plantarum*, 49, 355–359.
- Watanabe, S. (2010). Analysis equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., ... Wang, G. (2012). Field-based phenomics for plant genetics research. *Field Crops Research*, 133, 101–112.
- Yi, Q., Xiang, B., Ming, A., Wang, Q., & Zhao, J. (2013). Estimation of leaf water content in cotton by means of hyperspectral indices. *Computers and Electronics in Agriculture*, 90, 144–151.